A Temporal Analysis of Persuasion Principles in Phishing Emails

Olga A. Zielinska, Allaire K. Welk, Christopher B. Mayhorn, Emerson Murphy-Hill
North Carolina State University

Eight hundred eighty-seven phishing emails from Arizona State University, Brown University, and Cornell University were assessed by two reviewers for Cialdini's six principles of persuasion: authority, social proof, liking/similarity, commitment/consistency, scarcity, and reciprocation. A correlational analysis of email characteristics by year revealed that the persuasion principles of commitment/consistency and scarcity have increased over time, while the principles of reciprocation and social proof have decreased over time. Authority and liking/similarity revealed mixed results with certain characteristics increasing and others decreasing. Results from this study can inform user training of phishing emails and help cybersecurity software to become more effective.

INTRODUCTION

Phishing has been defined as "a criminal mechanism employing both social engineering and technical subterfuge to steal consumers' personal identity data and financial account credentials." (APWG, 2015). The Anti-Phishing Working Group elaborates on social engineering as "the use of emails posing as legitimate businesses that lead users to divulge information". This information could include usernames, passwords, dates of birth, social security numbers, or credit card numbers, among other things (Zielinska, Tembe, Hong, Ge, Murphy-Hill, & Mayhorn, 2014).

Multiple studies have assessed the end user's ability to accurately identify phishing emails, personality and demographic characteristics that could increase susceptibility to phishing, and training programs to better equip users against phishing attacks (Hong, Kelley, Tembe, Murphy-Hill, & Mayhorn, 2013; Sheng, Holbrook, Kumaraguru, Cranor, & Downs, 2010; Welk, Hong, Zielinska, Tembe, Murphy-Hill, & Mayhorn, 2015; Zielinska et al., 2014). Although these studies are useful in learning more about users' ability to identify phishing emails, users are still falling victim to phishing emails. In the first three quarters of 2015 alone, over 1 million unique phishing campaigns have been identified (APGW, 2015). As phishing emails continue to infiltrate users' mailboxes, perhaps the research focus should shift from the user to the email. Specifically, what social engineering techniques are the phishers using to successfully persuade victims into releasing sensitive information?

The purpose of this study is to examine persuasion principles as they are used in the context of phishing emails.

Related Work

Cialdini (2007) introduced six principles of influence that examined the Psychology of Persuasion. These six principles (authority, social proof, liking/similarity, commitment/consistency, scarcity, and reciprocation) have been linked to elements of phishing emails (Akbar, 2004; Ferreira, Coventry, & Lenzini, 2015). Below is a brief description of each principle and an example of how each might be used in a phishing attempt.

Authority. Individuals tend not to question authority. This could be out of fear to avoid negative consequences such as losing privileges, humiliation, or condemnation. Authority could be displayed as an official signature/logo or an email coming from an administrator (Akbar, 2004; Cialdini, 2007).

Social Proof. People will let their guard down if they believe everyone around them shares the same risk. They also want to be included in what other people are doing. Social proof could include an email referring to other customers like you that viewed this item with a URL leading you to the page (Akbar, 2004; Uebelacker & Queil, 2014).

Liking/Similarity. We tend to be easily persuaded by people we know and like or people who are similar to ourselves. Additionally, people trust those they find attractive or credible, and trust increases compliance. If an email is similar to previous emails from an organization or appears credible, users could trust the message and comply with requests (Akbar, 2014; Cialdini, 2007; Uebelacker & Queil 2014; Workman, 2008).

Commitment/Consistency. Users will honor commitments they have previously made and be consistent with their actions. If a company reminds a user of the terms of use and their agreement to change their password yearly, they may feel committed to these terms feel obligated to follow the request (Akbar, 2014).

Scarcity. An emotional response is elicited when the availability of an item/service is limited or there is only a short time frame to respond. Hackers could include a threatening message such as: respond to this email within 24 hours or you lose access to your email account, money in your bank account, or a host of other consequences (Akbar, 2014; Cialdini, 2007, Uebelacker & Queil, 2014).

Reciprocation. This social norm obligates individuals to repay others for a service they have received. A user may receive an email from a company indicating that their account had suspicious activity. They may appreciate this notification and reciprocate with updating their account information to prevent this from happening again (Akbar, 2014; Cialdini, 2007; Uebelacker & Queil, 2014; Workman, 2008).

Akbar (2014) analyzed 207 phishing emails obtained from a Netherlands database over a 4-month period based on Cialdini's principles. He found that the authority and scarcity were the most common persuasion principles used in phishing emails. Ferreira, Coventry, and Lenzini (2015) performed a similar analysis on 52 phishing emails obtained from their own inboxes and common phishing emails posted online. Ferreira et al. (2015) combined persuasion principles with Gragg's psychological triggers, and Stajano's principles of scams. The results of Ferreira et al's (2015) analysis were slightly conflicting with those of Akbar (2014) with liking/similarity as the most common principle, followed by scarcity, and then authority as the third most popular persuasion principle.

These studies provide a strong theoretical background and preliminary results of persuasion principles present in phishing emails; however, there are a few limitations to consider. First, each study was limited in their sources of phishing emails. Emails included in the Akbar (2014) study were retrieved from one national database and the emails in the Ferrira et al. (2015) study were from the researchers' own email accounts. This could bias the sample selection and the results they found. It would be beneficial to analyze emails from more than one available source.

Additionally, the emails analyzed were both from international sources, specifically, the Netherlands, Luxembourg, and England. Previous research has revealed differences in the perceptions of phishing among American, Chinese, and Indian cultures (Tembe, Zielinska, Liu, Hong, Murphy-Hill, & Mayhorn, 2014). There may also be cultural differences in phishing emails used between European countries and the United States. Analyzing emails targeting users in the United States could offer a different perspective to determine if phishers are utilizing similar techniques or if they have a unique approach when targeting users in the United States. Finally, the emails chosen from previous research had a small range of dates, typically a few months. Little is known about the trends of the social engineering principles in phishing emails over time. Looking back at the history and the evolution of persuasion principles in phishing emails can reveal tactics hackers may have used in the past, what the most popular current items are, and possibly predict what elements will be used in the future.

METHOD

Materials

Eight hundred eighty-seven emails were included for this study. We retrieved these emails from three sources: Brown University, Cornell University, and Arizona State University. Each of these universities provided examples of phishing emails that have circulated on their campus. The examples are updated on a regular basis. They also provided an archive of emails that dated back to 2010 (Arizona (2011-2015; https://getprotected.asu.edu/phishing); Brown (2014-2015; https://it.brown.edu/alerts/phishing); Cornell (2010-2015; https://www.it.cornell.edu/security/phishbowl.cfm). For this study, all emails available from these three sources through June 11, 2015 were used. Each email was saved as a PDF file and given a unique letter and number identifier. Emails were saved in the event that the website would be disabled or changed during the duration of the study.

Each email was evaluated using a questionnaire adapted from the Ferreira et al. (2015) study. The information collected from each email is listed below. The persuasion principle/principles are identified with each question. Persuasion principles are not mutually exclusive and certain questions may assess one or more persuasion principle. If a persuasion principle is not identified with a question, the information was collected for tracking purposes. Additionally, dates were collected to track persuasion principles over time. Potential responses for each question are listed below their corresponding question. If the response "other" was selected, reviewers had the opportunity to enter more details in a free response text box.

- Email ID
- Date Listed
- Attachment Present
    - Yes
    - No
- Link Present
    - Yes
    - No
- Reply Requested (Reciprocation)
    - Yes
    - No
- Is the email from the following: (Authority, Liking/Similarity)
    - Government
    - Educational Institution
    - Banking Agency
    - Other
    - None
- Is there a logo? (Authority, Liking Similarity)
    - Yes
    - No
- Is the email asking the user to perform an action? (Authority, Consistency/Commitment, Reciprocity)
    - Click here/ Click link/ "Click"
    - Update form
    - Confirm form
    - Open the attachment
    - Confirm personal information
    - Upgrade account information
    - Other action asked to be performed
    - None
- Does the contain information regarding known contacts? (Social Proof, Liking/Similarity)?
    - Friends
    - Colleagues
    - Family
    - Other information regarding known contacts
    - None
- Does the email refer to actions performed by other users? (Social Proof, Liking/Similarity)

  - Customer complaints
  - Others expecting your input
  - Other actions performed by other users
  - None
- Does the email contain the following identifying information? (Liking/Similarity, Authority)
  - Email address
  - Physical address
  - Telephone number
  - Other identifying information
  - None
- Are there details included in the email? (Liking/Similarity)
  - Invoice number
  - Requested service details
  - Payment details
  - Other details of service
  - None
- Are there elements in the first person stating "I am this or that"? (Liking/ Similarity)
  - Yes
  - No
- Are there elements in the first person describing the behavior around others? (Social Proof, Liking/Similarity)
  - Yes
  - No
- Is the email referring to other elements outside the email to look more reliable (Adobe Reader, etc.)? (Liking/Similarity, Consistency/Commitment)
  - Yes
  - No
- Is the email asking commitment from the user? (Commitment/Consistency, Reciprocation)
  - "Can I trust you?"
  - "Can you do this for me?"
  - Other commitments
  - None
- Does the email have visual cues? (Liking/Similarity)
  - Colors
  - Unusual font
  - Abnormal use of capital letters
  - Big images
  - Exclamation and/or interrogation marks
  - Spelling mistakes
  - Grammar mistakes
  - Other visual cues
  - None
- Does the email convey a sense of urgency? (Scarcity)
  - Time restrictions
  - "Urgent"
  - "Must be done"
  - Other items conveying sense of urgency
  - None
- Does the email list a consequence if user does not comply? (Authority)
  - Yes
  - No

Two reviewers assessed the emails independently using the questionnaire and recorded their results in Qualtrics, a survey collection system. There was an average agreement of 87% per item between the two raters.

Analysis

Once each email was assessed results were exported and analyzed in SPSS. The frequency of each response was calculated and presented in the next section. A Spearman's Rho correlational analysis was conducted to compare each phishing element over time. Spearman's Rho is robust against non-normally distributed data. As seen in the next section, there was an uneven distribution of emails over time, therefore, Spearman's Rho was used to compare phishing elements over time.

RESULTS

The results section is divided into two sections. The first section assesses the frequency of each item response. The second section contains the results of the correlational analysis of email characteristics over time.

Frequency of Email Characteristics

The majority of the emails (96%) identified the year the phishing email was circulating. Figure 1, below, shows the distribution of the emails by year. It is important to note that emails were only collected from January through June for the year 2015.

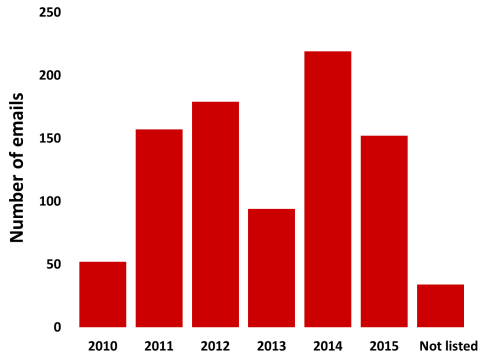

Figure 1: Number of phishing emails available by year

Of the 887 emails, the majority included a link leading to a phishing website (80%). A reply request was the next most common feature (11%), followed by an attachment (10%). Seventeen emails included two or all three of the following: link, attachment, reply requested.

Hackers posed as many different sources to trick users. The most common was an educational institute (57%). The second most selected option was "Other" (33%). Of the 33% of "Other" responses, 85% posed as companies. Analyzing the free text responses, companies such as Google, Apple, Microsoft, and Paypal were among the top spoofed in phishing emails. Government and financial institutions each accounted for 8% of the email sources.

Logos were present in only 5% of emails. When assessing what actions were requested from the user, several variations of "Click" such as "Click Here" or "Click Link" appeared in over half of the emails (54%). Twenty-one percent of responses were indicated in the "Other" category. The "Other" category revealed that hackers asked users to follow/visit/use the link 68 times, contact/send an email/reply 30 times, confirm/verify/authenticate information 30 times, and to download an item 24 times. These requested actions closely follow the frequency of the presence of links, the presence of attachments, and the number of times a reply was requested.

The majority of emails did not contain information regarding known contacts (98%), actions performed by other users (89%), identifying information such as email or physical address (85%), details of service (85%), elements in the first person (98%), elements describing the behavior of others (99%), elements referring outside the email to look more reliable (95%), or asking commitment from the user (97%).

Conversely, a large portion of emails contained visual cues (95%). The most common visual cues included grammar mistakes (81%), followed by abnormal uses of capital letters (64%), and unusual fonts (31%). Unlike other email characteristics, the presence of certain visual cues could lower the credibility of the email, in turn lowering the trust toward the emails, and consequently become less persuasive. For example if an email contains multiple grammar errors and abnormal uses of capital letters, the user may become suspicious and focus more on the visual cues than the content of the message.

Urgent elements were present in 34% of the emails. These could include time restrictions (15%), such as requiring an action to be completed within 24 hours, indicating that this is a "Final Warning", or that an action is required. Finally, a consequence for failing to follow the instructions of the email was contained in 34% of the emails. The most common consequence listed was account suspension or termination (75%) followed by the inability to send or received emails (13%).

Temporal Analysis

Spearman's Rho correlations were used to compare email characteristics over time. During the five year period under consideration (2010-2015), the presence of a link to a phishing website rose ($r_s(850) = .10$, $p = .004$), while the request for a reply decreased ($r_s(850) = -.12$, $p = .001$). There was also a shift in the source of the email. Specifically, the number of emails that were posed as coming from an educational institute increased ($r_s(850) = .21$, $p < .001$), while emails coming from financial institutes ($r_s(850) = -.18$, $p < .001$) and other sources, such as companies, decreased significantly ($r_s(850) = -.15$, $p < .001$). The shift in sources of the emails could indicate that hackers are learning more about their potential victims and targeting their emails to be from sources they are more likely to open. It is not unlikely that students will receive an email from their own university; however, it could be suspicious if they receive an email from a bank or company they have not used in the past.

There was also an increase in the number of emails that contained a logo ($r_s(850) = .18$, $p < .001$). Although only 5% of emails contained logos in our analysis, this could be a trend that is increasing due to an intent to raise the credibility and authentic appearance of an email. Additionally, emails referring to actions performed by other users is also decreasing ($r_s(850) = -.10$, $p = .01$). It may be that hackers have found this tactic less successful or less prevalent in authentic emails, therefore their use has decreased over time. Additionally, service details have decreased over time ($r_s(850) = .13$, $p < .001$), especially payment details ($r_s(850) = -.16$, $p < .001$).

An increase in referring to other elements outside the email to look more reliable, such as Google Docs or Adobe Reader, has also been identified ($r_s(850) = .12$, $p = .001$). Additionally, visual cues such as colors ($r_s(850) = .18$, $p < .001$), unusual fonts ($r_s(850) = .25$, $p < .001$), big images ($r_s(850) = .12$, $p = .001$), and grammar mistakes ($r_s(850) = .09$, $p = .007$) have increased, while exclamation marks ($r_s(850) = -.11$, $p = .002$) and other visual cues have decreased ($r_s(850) = -.16$, $p < .001$).

Lastly, there was an increase in the number of emails that contained elements of urgency ($r_s(850) = .09$, $p = .01$). These elements could encourage users to act quickly and may have success in eliciting a response from users, which is why it is on the rise. A table summarizing the increase and decrease of email characteristics, along with their corresponding persuasion principles are listed in Table 1.

Table 1: Email characteristics that increased or decreased over time, their correlational value, and the associated persuasion elements.

Increase		**Persuasion Element**
Link	.10**	
Source *Educational Institute*	.21***	Authority, Liking/Similarity
Logo	.18***	Authority, Liking/Similarity
Referring to elements outside the email	.12**	Commitment/Consistency, Liking/Similarity
Visual Cues *Colors* *Unusual Font* *Big Images* *Grammar Mistakes*	 .18*** .25*** .12** .09**	 Liking/Similarity
Urgency	.09*	Scarcity
Decrease		**Persuasion Element**
Requested Reply	-.12**	Reciprocation
Source *Financial Institute* *Other*	 -.18*** -.15***	Authority, Liking/Similarity
Actions performed by other users	-.10*	Social Proof, Liking/Similarity
Service Details *None,* *Payment Details*	 .13*** -.16***	 Liking/Similarity
Visual Cues *Exclamation Mark* *Other Visual Cues*	 -.11** -.16***	 Liking/Similarity

*p<.05 **p<.01 ***p<.001

DISCUSSION

Phishers implement a variety of strategies to hook their victims and persuade them into revealing sensitive information. Emails could be manipulated to make them look authoritative, urgent, and similar to previous emails they have encountered. Through our analysis, we have found that most phishing emails contain links to guide users to fake websites where they can enter their information. These emails could include logos and references to outside elements (i.e. Google Docs, Adobe Reader) to make them look more credible and consistent. Additionally, they could be from known sources to increase trustworthiness. For example, the majority of the emails to educational institutions were staged to be from the university.

There were also changes in the email characteristics over time. Specifically, we saw an increase in the persuasion principles of commitment/consistency and scarcity over time. This could be seen through an increase in referencing elements outside the email to maintain consistency with legitimate emails and an increase of time restrictions to increase urgency. There was a decrease over time in the persuasion principles of reciprocation and social proof. Examples of this include the decrease of requested replies in emails over time and the decrease of references of other user's actions in emails.

Two persuasion principles exhibited both an increase and decrease in their presence in emails over time: authority and liking/similarity. These principles could increase phishing rate success if used appropriately, but could also raise suspicions in users and decrease compliance if used incorrectly. One example could be seen through the source of the email. Emails in this study were pulled from three academic universities. Phishers may have had more success when they posed emails from the same academic institution where targeted students attended. This would not raise suspicion. If the phishers tried to pose as a student's bank, for example Bank of America, and the student is a customer of Wells Fargo, this could raise suspicion to the student and they would not fall prey to the phish. This could explain the trend of increasing educational institution sources and the decrease in banking agencies as the source.

The results from this study offer a different perspective regarding phishing. Previous research has focused on the user aspect; however, few studies have examined the phisher perspective and the social psychological techniques they are implementing. Additionally, they have yet to look at the success of the social psychology techniques. Results from this study can be used to help to predict future trends and inform training programs, as well as machine learning programs used to identify phishing messages.

There are a few limitations that need to be considered with this study. First, the emails assessed in this study were taken from publicly posted phishing email websites at three academic universities. This may not include the full set of phishing emails that are in circulation, as many phishing emails are not reported. Also, there may be phishing emails that are still undetected and therefore not reported. Furthermore, the education domain is only one of many targeted in phishing. It would be beneficial to assess phishing emails targeted toward government institutions, financial institutions, and personal email accounts, among many others to understand the social engineering of phishing emails more thoroughly.

Finally it would be interesting to combine the research of user characteristics and email characteristics to examine if there is a relationship. It is an empirical question if certain persuasion theories would be more or less beneficial based on user personality traits. If a relationship is determined, training could be targeted to individual differences to target any vulnerabilities. Additionally, phishing detection systems could be specialized to each user to compensate for particular user attributes that might be exploited by certain attacks.

ACKNOWLEDGEMENTS

The authors would like to thank the Science of Security Lablet for funding this project. The authors would also like to thank Andrea Zani and Phoebe Pradhan for their assistance in data coding.

REFERENCES

Akbar, N (2014). *Analysing Persuasion Principles in Phishing Emails.* Unpublished Master's Thesis. University of Twente. Enschede, Netherlands.

Anti-Phishing Working Group (APWG) (2015). Phishing Activity Trends Report http://docs.apwg.org/reports/apwg_trends_report_q1-q3_2015.pdf

Cialdini, R. B. (2007). *Influence: The psychology of persuasion*. New York: Collins.

Ferreira, A., & Lenzini, G. (2015, July). An analysis of social engineering principles in effective phishing. In *Socio-Technical Aspects in Security and Trust (STAST), 2015 Workshop on* (pp. 9-16). IEEE.

Hong, K. W., Kelley, C. M., Tembe, R., Murphy-Hill, E., & Mayhorn, C. B. (2013). Keeping Up With The Joneses Assessing Phishing Susceptibility in an Email Task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 57, No. 1, pp. 1012-1016). SAGE Publications.

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F., & Downs, J. (2010). Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 373-382). ACM.

Tembe, R., Zielinska, O., Liu, Y., Hong, K. W., Murphy-Hill, E., Mayhorn, C., & Ge, X. (2014, April). Phishing in international waters: exploring cross-national differences in phishing conceptualizations between Chinese, Indian and American samples. In *Proceedings of the 2014 Symposium and Bootcamp on the Science of Security* (p. 8). ACM.

Uebelacker, S., & Quiel, S. (2014, July). The social engineering personality framework. In *Socio-Technical Aspects in Security and Trust (STAST), 2014 Workshop on* (pp. 24-30). IEEE.

Welk, Hong, Zielinska, Tembe, Murphy-Hill, Mayhorn (2015). Will the "Phisher-Men" Reel You In? Assessing Individual Differences in a Phishing Detection Task. *International Journal of Cyber Behavior, Psychology and Learning*. 5(4), 1-16.

Workman, M. (2008). Wisecrackers: A theory-grounded investigation of phishing and pretext social engineering threats to information security.*Journal of the American Society for Information Science and Technology*,*59*(4), 662-674.

Zielinska, O. A., Tembe, R., Hong, K. W., Ge, X., Murphy-Hill, E., & Mayhorn, C. B. (2014). One Phish, Two Phish, How to Avoid the Internet Phish Analysis of Training Strategies to Detect Phishing Emails. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 1466-1470). SAGE Publications.